Journal of Theoretical Biology 462 (2019) 304–310

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/jtb

Moderate sex between protocells can balance between a decrease in assortment load and an increase in parasite spread

Zsuzsanna Vig-Milkovics[a], István Zachar[b,c,d,*], Ádám Kun[b,c,d], András Szilágyi[b,c,d], Eörs Szathmáry[a,b,c]

[a] Department of Plant Systematics, Ecology and Theoretical Biology, Eötvös Loránd University, 1117 Budapest, Hungary
[b] Evolutionary Systems Research Group, Centre for Ecological Research, Hungarian Academy of Sciences, 8237 Tihany, Hungary
[c] Center for the Conceptual Foundations of Science, Parmenides Foundation, 82049 Pullach/Munich, Germany
[d] MTA-ELTE Theoretical Biology and Evolutionary Ecology Research Group, 1117 Budapest, Hungary

ARTICLE INFO

Article history:
Received 23 May 2018
Revised 17 November 2018
Accepted 20 November 2018
Available online 22 November 2018

ABSTRACT

Sexual reproduction is widespread in nature despite the different kinds of cost that it entails. We do not know exactly when the first sexual process took place and especially why it was beneficial at first. It is clearer why sex is advantageous for the prokaryotes and eukaryotes but the benefit of sex for protocells with individually replicating ribozymes is not yet fully understood. In this context sex is the simple horizontal gene transfer among two protocells that undergo transient fusion. Many authors argue that horizontal gene transfer (HGT) was very common in the early stage of evolution. However, HGT is a risky mechanism considering both the disruption of optimal compositions and the spread of parasites among protocells. In order to test the effects of HGT on the fitness of a protocell population, we explored by numerical simulations those conditions under which fusion might have been beneficial. We investigated multiple conceivable types of fusion in the stochastic corrector model framework and we considered the spread of parasites in every case. Protocells contain up to five species of unlinked, essential ribozymes; if a protocell has the same amount of each, it reaches maximum fitness. Fusion is dangerous not only due to the spread of parasites but also because it can ruin the cells with balanced ribozyme composition. We show that fusion can restore the ribozyme composition of the protocells under certain circumstances (high gene count, intermediate split size and low rate of fusion) and thus it can decrease the effect of the genetic load. Fusion could have been a useful early mechanism in contributing to the reliable coexistence of the different ribozymes before the spread of the chromosomes.

1. Introduction

The evolutionary path leading from mere organic molecules to modern cells, through the RNA world (Joyce, 2002; Kun et al., 2015; Yarus, 2011), witnessed tremendous increase in complexity, which requires considerable information-integration capacity. But erroneous replication limits the amount of information that can be passed on (Eigen, 1971). This size-limit was calculated to be about 100 nucleotides, which might be enough to store information for a single ribozyme (RNA enzyme), but definitely less than required for a living organism. Even this rather bleak conclusion assumes that the replication of the master sequence is faster than that of its mutants. Sol Spiegelman's experiment (1967) has demonstrated that short sequences are selected for, and the functional sequence is quickly outcompeted in a well-mixed system. There is one way to allow the master sequence to coexist with its mutants: the introduction of a higher level of selection. The encapsulation of hereditary replicators to form protocells is a viable solution in experimental setups (Matsumura et al., 2016).

Protocells (prebiotic reproducing compartments enclosing some genes and, often, metabolites) are assumed to have been an important interim phase of early evolution (Rasmussen et al., 2008; Schrum et al., 2010; Szathmáry et al., 2005). The origin of protocells encapsulating hereditary replicators are regarded as the first major evolutionary transitions (Maynard Smith and Szathmáry, 1995; Szathmáry, 2015; Szathmáry and Maynard Smith, 1995) of the egalitarian type (Queller, 1997). It is a transition in individuality (Michod, 1999) as well as in hereditary information trans-

* Corresponding author at: EVOLGINOP Research Group, Klebelsberg Kuno u. 3, 8237 Tihany, Hungary.
E-mail address: istvan.zachar80@gmail.com (I. Zachar).

https://doi.org/10.1016/j.jtbi.2018.11.020

fer, as it can also be regarded as the emergence of the genome, as opposed to the dispersal dynamics of naked, unlinked replicators. However, since protocells encapsulated unlinked, thus very likely competing, replicators, this can be regarded as the first appearance of intragenomic conflict as well.

Note that it is quite likely that the protocells followed the replication of "naked" genes, spreading on surfaces: for recent reviews of those models see (Czárán et al., 2015; Scheuring et al., 2003; Szilágyi et al., 2017; Takeuchi and Hogeweg, 2012). Under those circumstances different genes were mixing continuously in the population, but there is enough viscosity to the system to allow the coexistence of replicators. Although passive compartmentation can ensure coexistence of a limited number of genes (Boerlijst and Hogeweg, 1991; Könnyű and Czárán, 2013), active compartmentation in the form of reproducing compartments, possibly at a later evolutionary stage, is a more powerful coexistential mechanism.

The stochastic corrector model (SCM; Grey et al., 1995; Szathmáry and Demeter, 1987) was conceived to describe the process of information integration (Szathmáry, 1989) at the protocell level, despite internal competition at the replicator (gene) level. It is a dynamic and (in some realizations) a fully analytic model of group selection of early replicators. The variation on which natural selection at the protocell level acts is due to two types of stochasticity: first, there is demographic stochasticity in replicator growth (due to low numbers, deterministic chemical kinetics of replication does not apply), and second, there is stochasticity in assortment of the unlinked replicators to the offspring compartments. The SCM operates at a balance of opposing evolutionary forces. There is a dosage effect because the replicators are assumed to exert enzymatic function also. This by itself selects for high copy numbers, as does the need to decrease the assortment load, understood as the chance loss essential of replicators upon stochastic protocells fission. According to a recent study of the SCM where mutations were entirely omitted to focus on the information integration aspect, up to 100 genes can coexist in a protocell (Hubai and Kun, 2016) if redundancy (the number of copies) is high enough. But too high copy numbers come at a price: group selection loses its force due to decreased stochasticity, metabolic costs increase, the mutational load increases and the effect of new positive mutants is diluted out. Early genomes must have operated between "Scylla and Charybdis" in several dimensions (Koch, 1984; Niesert et al., 1981).

The assortment load can be decreased by two mechanisms: the spread of chromosomes (Maynard Smith and Szathmáry, 1993) and protocellular sex (Bernstein et al., 1984). We do not yet know the quantitative conditions of the first mechanism since the effect of deleterious mutations – in contrast to the investigations of Maynard Smith and Szathmáry (1993) – remain to be taken into account in a forthcoming model). While considering the neutral mutations allowed by the genotype-phonotype map of ribozymes (Kun et al., 2005; Szilágyi et al., 2014; Takeuchi et al., 2005) and the slowing down of replication after the erroneous insertion of a nucleotide (Leu et al., 2012; Rajamani et al., 2010) allow for a somewhat relaxed error threshold, replication of the whole genome is still out of reach. There is one active mechanism that could further relax the error threshold: intragenomic recombination (Santos et al., 2004). Thus, there is an indication that protocellular sex can facilitate information integration.

Sexual reproduction is widespread in the nature despite the fitness advantage of asexual populations (Maynard Smith, 1978). We do not know exactly when the first sexual process took place and especially why it was beneficial at first. It is clearer why sex is advantageous for the prokaryotes and eukaryotes but the benefit of the sexual mechanisms for protocells with individually replicating replicators is not yet fully understood. In this context, sex is horizontal gene transfer, the exchange of unlinked genes among two protocells that undergo transient fusion (Santos et al., 2003). Many authors argue that horizontal gene transfer (HGT) was very common in the early stage of evolution (Poole, 2009; Vetsigian et al., 2006; Vogan and Higgs, 2011). Protocellular sex can indeed restore the whole genome complement, but Bernstein et al. (1984) did not consider the problem of intragenomic conflict, since they thought to have it resolved by the hypercyclic coupling among the replicators. The snag is that under high mutation rates the SCM performs better than the compartmentalized hypercycle (Zintzaras et al., 2002), thus we are back to intragenomic conflict. Lateral gene transfer of unlinked genes is dangerous because of the threat of parasites spreading (cf. the problem of biparental inheritance of defective cell organelles, (Maynard Smith and Szathmáry, 1995)) to other protocells.

Here we explore the conditions under which protocellular sex might have been beneficial, increasing protocell fitness. The closest precedent to our work is the investigation by Santos et al. (2003) of the spread of selfish replicase molecules with or without sex in the SCM. The authors considered protocells harbouring three types of gene: a replicase and two different metabolic genes, and considered only selfish and non-selfish replicases. Differences to the usual modelling practices for SCM dynamics included: (*i*) Metabolic templates were growing at approximately the same rate., (*ii*) lack of dosage effect by the metabolic genes (i.e. cells having at least one metabolic gene of both types had maximal fitness, and (*iii*) no intragenomic conflict between the two metabolic genes. In the case of deleterious recurrent mutations of the metabolic templates the fitness function used by Santos et al. (2003) was the same as in Zintzaras et al. (2002): multiplicative for the different ribozymes (note that, unlike in the present paper, there was no resulting dosage effect). Santos et al. (2003) found that without recurrent deleterious mutations, the spread of selfish replicases through occasional protocell fusion was harmful, but also that mutations of metabolic genes under high mutation rates accumulated more slowly in the presence of selfish replicases and could lead to increased average protocell fitness (due to fewer replication rounds of the metabolic genes).

In this paper, we take the converse route by assuming that replication just happens (as in the original SCM model) and that metabolic genes suffer from deleterious recurrent mutations (irreversibly losing metabolic function and replication ability) and that they can also mutate into purely selfish parasites. We set identical wild-type replication rates for the different types of metabolic genes in order to focus on the competition against parasites. Our main goal is to study the effect of the rate of fusion-fission cycles on the average fitness of the population under these assumptions.

2. Methods: the good, the bad and the ugly

Our model is based on the implementation of the stochastic corrector model (Grey et al., 1995; Szathmáry and Demeter, 1987). SCM describes the group selection dynamics of compartmentalized replicators. Stochasticity in replication and during cell division generates variation, on which natural selection – between the cells – can act, favouring compartments with a replicator composition closer to optimal. Taking this model as a starting point, we introduce horizontal gene transfer to the evolution of protocells. The mechanism of HGT is simply a transient fusion between protocells. We suppose that primitive protocells acquired the ability of fusion in the early stage of their evolution.

Inside a protocell, replicators replicate individually. We differentiate among three types of RNA molecules: (1) enzymatic ribozymes, which both replicate and contribute to the metabolism of the protocell; (2) inactive RNA fragments (junk) which lack the capacity to replicate; and (3) parasites which do not contribute to

the metabolism but can replicate. The considered molecular types have different replication rates (r) and different enzymatic activity (E).

Ribozymes belong to one of N ribozyme classes (up to five) which form a chain of reactions ($E > 0$). Each enzymatic ribozyme can be in one of the two forms: a metabolically efficient but slowly replicating good form (denoted by the index g) and an inefficient but fast replicating form (denoted by the index b). Thus, there is a trade-off between the affinity to the replicase and ribozyme activity, i.e. the higher the ribozyme activity, the lower the ribozymes' affinity to the replicase. Bad ribozymes therefore replicate faster than good ones ($r_g < r_b$), but they have lower enzymatic activity ($E_g > E_b$). The enzymatic activity of the good (and the bad) state is the same per ribozyme species; $E_{g_i} = E_g = 1$ (and $E_{b_i} = E_b = 0.2$, with $i = 1, 2, \ldots N$). The replication rates are assumed to be the same for the different types of metabolic gene ($r_{g_i} = r_g = 1$ and $r_{b_i} = r_b = 1.1$ for $i = 1, 2, \ldots N$), so there is no competition between the wildtype ribozyme species. Parasites have the highest replication rate, higher than that of bad ribozymes ($r_g < r_b < r_p = 1.2$), and they have no enzymatic activity ($E_p = 0$). Junk RNA fragments have neither enzymatic activity ($E_{junk} = 0$), nor can they replicate ($r_{junk} = 0$). During initialization, the enzyme composition of each protocell is randomly drawn from a uniform distribution over $(0 \ldots \max_g)$ where $\max_g = 2$ for each of the N classes of good ribozymes; initially there is no bad ribozyme, junk or parasite within protocells.

Ribozymes catalyse steps in a series of reactions which is assumed to produce the material for RNA replication and protocell growth. We assume a constant I_0 material influx for the first reaction step. The transition of chemical entities by ribozymes from $(i-1)$th step to the ith step is catalysed by the good (g_i)or the bad (b_i)state of a given ribozyme. The flux of the ith elementary reaction step is

$$I_i = I_{i-1}\left(n_{g_i}E_{g_i} + n_{b_i}E_{b_i}\right), \tag{1}$$

where n_{g_i} and n_{b_i} are the number of ribozymes of the ith species with high and low activity respectively. The final flux is

$$I_N = I_0 \prod_{i=1}^{N} \left(n_{g_i}E_{g_i} + n_{b_i}E_{b_i}\right). \tag{2}$$

Note the multiplicative effect on total flux. The final flux determines the rate of biomass accumulation. The replication rate of the cell depends on this flux; thus, it is considered as a measure of fitness. The final flux reaches the maximum value when the cell contains only the good states of each ribozyme class ($n_{b_i} = 0,\ i = 1, 2, \ldots N$) in a uniform distribution ($n_{g_i} = s/N,\ i = 1, 2, \ldots N$) and there is no parasite in the cell. The split size s is the predetermined "volume" (total amount of replicators, junk and parasite included) at which cells split, uniform to all cells. Therefore, the optimal flux is

$$I_N^{opt} = I_0\left(\frac{s}{N}E_g\right)^N. \tag{3}$$

The fitness is the actual flux normalized with the optimal flux:

$$w = \frac{I_N}{I_N^{opt}} = \frac{\prod_{i=1}^{N}\left(n_{g_i}E_g + n_{b_i}E_b\right)}{\left(\frac{s}{N}E_g\right)^N} = \frac{\prod_{i=1}^{N}\left(n_gE_g + n_bE_b\right)}{\left(\frac{s}{N}E_g\right)^N}. \tag{4}$$

Thus, the fitness of protocells is calculated based on the internal composition, and only the main flux is taken into account. The protocell is considered dead (having zero fitness) if it lacks both the good and bad states of any of the ribozyme classes, and thus the chain of reactions is disrupted. The proportion of dead cells in the population with five and four enzymes represents Figs. S1 and S2.

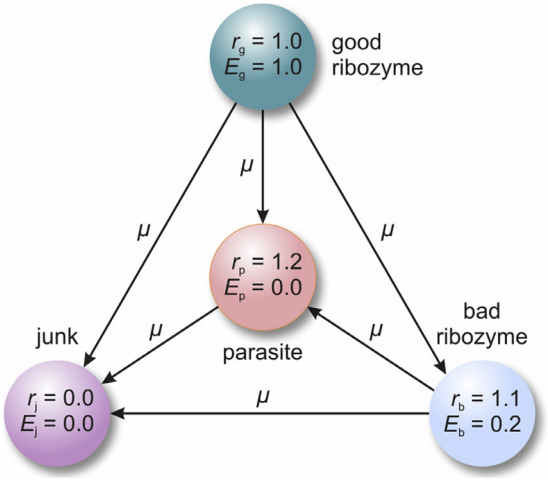

Fig. 1. Schematic representation of the ribozyme states within a ribozyme class and all the possible mutation paths between the ribozyme states. r_g, r_b, r_p, r_{junk} denote the replication rate of the ribozyme states good, bad, parasite and junk, respectively; and E_g, E_b, E_p, E_{junk} denote the enzymatic affinity of the ribozyme states. μ denotes the mutation rate between two ribozyme states as the probability of transition according to the arrow; Back mutations are neglected in the model.

At each update of the system, we randomly select a cell with probability proportional to its fitness. For the selected cell, we randomly choose a replication reaction (ribozymes or parasites as junk do not replicate) with probability proportional to reaction propensities. Propensity is defined as $P_x = n_x r_x$ where x stands for the various molecular types (parasite p included) $x = \{g_1 \ldots g_N, b_1 \ldots b_N, p\}$ and r_i is the replication rate of the ith ribozyme (parasite included). During the replication of ribozymes, mutations can occur with a constant mutation rate (μ). Ribozymes can mutate from the good to the bad state and each ribozyme and the parasite can mutate to junk or parasite (Fig. 1). We exclude the possibility of reverse mutations. Mutation rate is assumed to be $\mu = 1\%$ in all the numerical investigations presented here.

Cells divide when the number of molecules (ribozymes, parasites, and junk) reach a critical value, the split size s. Split size is one of the key parameters of the system; we have explored split sizes between 20 and 100 in our simulations. At cell division, one of the daughter cells replaces the original cell and the other replaces a randomly chosen cell from the population. The population includes 1000 protocells. Accordingly, the population dynamics of protocells follows a Moran process, and the population size is kept constant. Replicators assort randomly to the daughter protocells, which is one of the sources of variation on which between-protocell selection can act.

We allow HGT from the initial step between cells with a given probability of fusion that is immediately followed by binary fission. During fission, molecules assort randomly into the two daughter cells. By this way, we effectively homogenize the composition of the fusion partners. At each time step, fusion happens with probability φ. See Fig. S3 for the actual measured average fusion and split rates when $\varphi = 10\%$. Two partners are chosen according to the fusion mechanism (random, directly or inversely proportional to fitness, under a given fitness threshold). Dead cells cannot be fusion partners. We have explored fusion rate from zero to 80% (see Fig. 4).

We considered multiple methods of protocell selection for fusion, namely random, directly proportional to fitness, inversely proportional to fitness, and, lastly, fusion under an arbitrarily chosen fitness threshold (in this case the fitness of both partners must be under this threshold). In the random case, we select individuals for fusion randomly, regardless of their fitness. In the directly

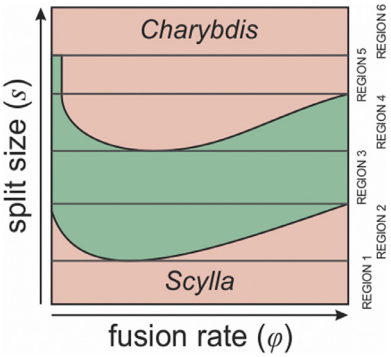

Fig. 2. Five parameter regions, in which the normalized fitness depends on the fusion rate. We can discern three parameter regions of split size if there is no protocellular sex. In one, degeneracy is too low, and thus the system dies out because of assortment load even without parasites. This is the Scylla of information integration: assortment load leads to loss of information which cannot be remedied by fusion. At high split sizes parasites have ample time to proliferate, and begin to dominate the population, causing dilution of the ribozymes. In effect the ribozyme population becomes low, and again because of assortment load, the protocell population becomes extinct. This is the Charybdis of information integration, when gene redundancy is too high and group selection loses its force. Here fusion does not help; on the contrary, it is detrimental to fitness. In-between the mythological monsters lie the traversable path where the population of the protocells is viable. At low split size, a small amount of protocellular sex can restore some protocells to viability. Hence there is Region 2, where without sex the population is not viable. This region is very narrow. Then there is Region 3 where the population is viable irrespective of mixing, but mixing can increase mean fitness (at around $s = 60$ in Fig. 3A). In Region 4, a small amount of mixing is disadvantageous, and viability is restored with high mixing. While at low split size, mixing can restore the gene content, at higher split size it also helps the spread of parasites. Mixing does not increase the overall parasite count but makes parasite load more even. This can be offset by higher split size, hence the shift of viability region toward higher split sizes. Region 4 is again narrow. In Region 5, protocellular sex is disadvantageous as it only helps the spread of parasites.

proportional case, cells with higher fitness are chosen for fusion with higher probability (Figs. S1d and S2d); in the inversely proportional case, cells with lower fitness are more likely to fuse. In case of fusion under a given fitness threshold, cells with favourable ribozyme compositions are excluded, so the composition of these cells does not deteriorate due to fusion. It is not worth fusing for cells with high fitness because their ribozyme composition is more balanced. Of course, it is not worth setting too low a threshold because fusion among cells with extremely low fitness does not lead to cells with a much-improved ribozyme composition.

3. Results

Protocellular sex can increase fitness of the population. With five metabolic genes and intermediate split size (s) a low rate ($\varphi = 0.05$) of fusion maximises fitness (Fig. 3A). Fitness at this point is roughly twice as high as at the no-fusion case. This is one of the outcomes, and we can distinguish five regions with increasing split size (Fig. 2).

Generally, we can observe the above five regions at different number of genes (Fig. 4), albeit some are more pronounced than others. At low gene count ($N = 2$) the negative effect of protocell fusion is pronounced: there is some benefit of sex at very low split sizes, but mostly it lowers fitness compared to no fusion. Assortment load is not a great concern for these systems, and thus the potential acquisition of missing genes from other cells cannot help much. At high gene count, on the other hand, the cell contains more ribozyme species in the same volume, consequently, it is easier to lose an essential ribozyme due to random splitting, so in this case acquiring a potentially useful ribozyme is the benefit to the cell. In this case, there is a chance that fusion of two protocells with disadvantageous ribozyme compositions will lead to more favourable daughter combinations after the random splitting. Reasonably, the disadvantageous composition means that there are many copies from one ribozyme class and a few from the rest.

In order to understand how fusion and reassortment can facilitate protocell survival, we have explored other methods of selecting partners for fusion. First, protocells were chosen for fusion proportional to their fitness. In this case, fusion was always disadvantageous (data not shown but see Figs. S1d and S2d). This is the most unfavourable condition because in this case the cells with good composition deteriorate due to the fusion and subsequent splitting. Second, protocells were chosen for fusion inversely proportional to their fitness, i.e. bad compositions were chosen with a higher probability. In this case, the maximum fitness increase that protocellular sex can achieve increased compared to the random fusion method (Fig. 4b). Thus, it is, as expected, the protocells with bad compositions that need to fuse and reassort their repli-

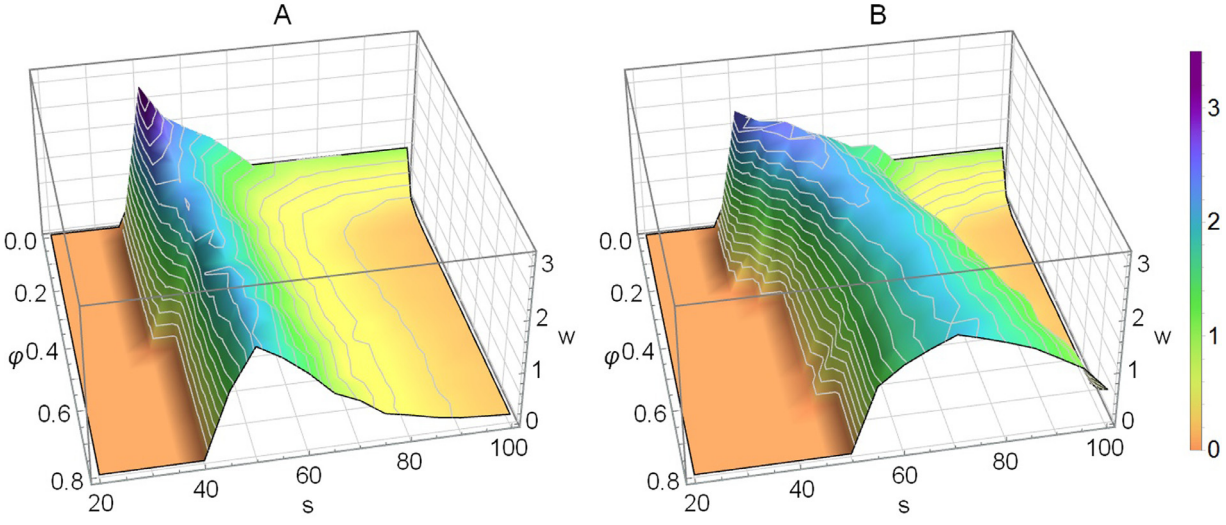

Fig. 3. The effects of fusion on the mean fitness of the protocells, when there are 5 ribozymes in the population. (**A**) random fusion; (**B**) fusion under fitness threshold $w = 0.03$. Axes x, y and z represent fusion rate φ, split size s and the mean fitness divided by the mean fitness in case of zero fusion, thus we get w_r, respectively. In the random case, the mean fitness is the highest when $\varphi = 5\%$. When fusion occurs only among the bad protocells (threshold $w = 0.03$) the fusion is beneficial when the fusion rate is very high, $\varphi = 75 - 80\%$. See Fig. 4 for all other potential cases. Colours are rescaled within the data range. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

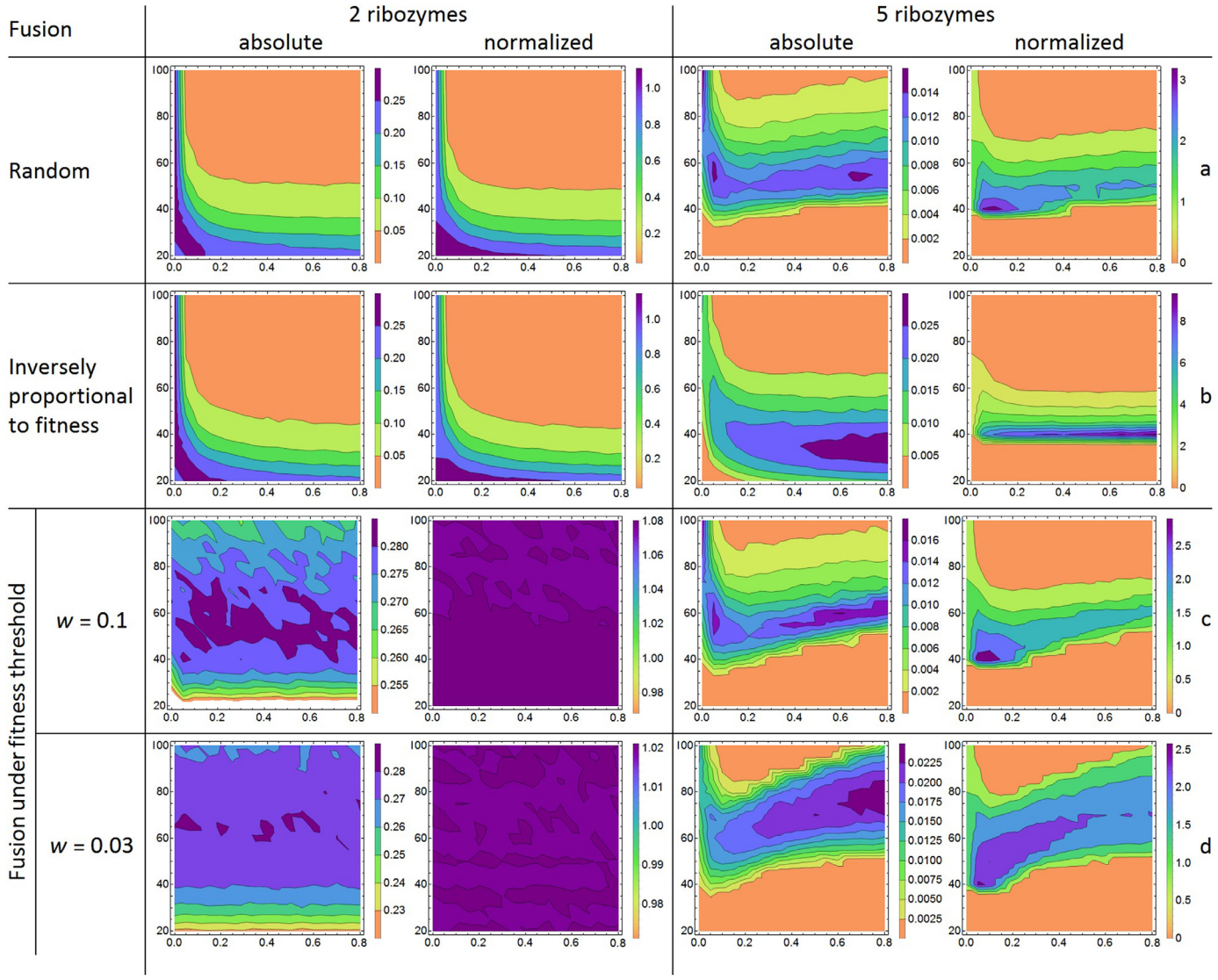

Fig. 4. The effects of the possible methods of fusion (except the directly proportional case, see Figs. S1d and S2d) on the mean fitness of the protocell population, when the protocells contain two and five ribozyme species. For all subplots, axes x and y represent fusion rate φ and split size s. Surfaces show the mean fitness of the population (**absolute**) or the normalized mean fitness divided by the mean fitness in case of zero fusion (**normalized**) as a function of fusion rate and split size. Purple/blue represents higher, orange lower ones (colours are scaled individually for each plot for the given z range, for emphasis). Methods of selection for fusion partners: random (**a**), inversely proportional to the fitness (**b**), fusion under an arbitrarily chosen fitness threshold (**c,d**) when the fitness of both cells is under this fitness threshold. Parameters: 10^7 time steps, 1000 protocells in the population, initial maximum for random good and bad are $n_g = 2$, $n_b = 0$ for each of their N classes, and there is no junk and parasite in the initial step. Actual value will be drawn randomly from $(0, N-1)$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

cators. With this method of choosing fusion partners, the increase in mean fitness shifts toward lower split size. The Charybdis part expands in size: big cells with bad compositions have a low probability to yield good compositions.

Finally, we assume cell fusion only below a fitness threshold (cells with favourable ribozyme compositions are excluded, hence fusion affects only the cells with markedly unfavourable ribozyme composition.). Fusion generally improves the mean fitness of the population (compare Fig. 3A and B). The region favouring fusion expands greatly. At low gene count, protocellular sex of any rate produced systems with equal or greater fitness compared to a system without fusion. Thus, limiting fusion to protocells having bad composition ensures that fusion has a chance to improve the composition. However, when the fitness of the fusion partners is too low (Fig. S4), fusion is less useful because two cells with unfavourable ribozyme compositions are less likely to produce a cell with a much better composition. In the latter case the increase of fusion rate has almost no effect; this is because only a small fraction of the population has such a small fitness, thus fusion events are rare.

To further examine the direct effects of sex on fitness, we performed a complementary investigation. We assumed four enzymatic template species and formed random compartments consisting of at most 80 templates (including high and low activity versions of ribozyme types, parasites and junk) and binned them according to their fitness values. The fitness of the compartment was computed according to Eq. (4). We have generated a total of 15,000 compartments, out of which 1000 compartments fell in the $[0.00, 0.02]$ fitness interval, 1000 fell in the $[0.02, 0.04]$ interval, ..., and 1000 compartments fell in the $[0.28, 0.3]$ interval. (Compartments with higher fitness are very rare.)

From these 15,000 compartments we chose pairs randomly, implementing fusion and fission by reassorting ribozymes according to the hypergeometric distribution, then picked one of the result-

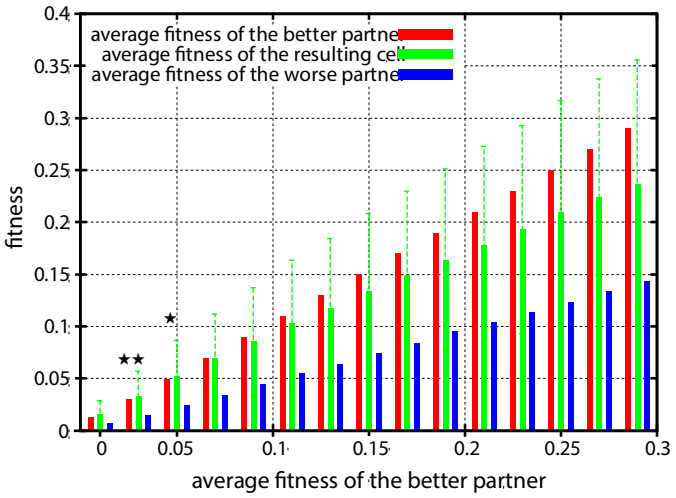

Fig. 5. Binned average and standard deviation of the fitness of the daughter cells (green bars) as a function of the better partner of fusion (red bar). The average of the worse partner is also presented (blue bars). The good and bad activity of ribozymes is 1.0 and 0.2 respectively. Each protocell consists of at most 80 ribozymes (including parasites and junk). The values in each bin are the average of 10^4 random fusions. Bins marked with asterisk can be compared to the results shown in Figure S4. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ing two cells randomly. We made 1.5×10^5 random pairings, and the fitness values of the selectively inferior partner and the resulting cell are plotted as the function of the fitness of the better partner, see (Fig. 5). It can be seen that sex can be useful on average when the selection threshold is $w<0.05$, in line with the individual-based model. The relatively high variance refers to the highly random nature of the reassortment. The same position of the fitness threshold of $w=0.05$ in the individual-based model and this simple model (despite to the fact the latter is an oversimplification, mainly because of taking only one round of sex into account) shows that the selection for the potentially emerging better compartments by sex does not operate above the fitness threshold.

4. Discussion

We explored by numerical simulation of the population dynamics the conditions under which fusion could give benefit to protocells and the mean fitness of the population could increase. Protocellular sex is not a universal remedy for the informational crisis these cells face. But there is a chance that the occasional fusion of two protocells with not overly disadvantageous ribozyme compositions will lead to more favourable combination of ribozymes after random splitting. There is a parameter range of the fusion rate, the split size and the fitness threshold of the cells selected for fusion, where HGT is beneficial (Fig. 4). Fusion can help the protocell with less favourable ribozyme composition to obtain useful ribozymes and thus avoid death, but fusion is damaging when it happens between cells with high fitness. Accordingly, we can conclude that cell fusion in the early evolution of life could have been advantageous under certain circumstances and it could have contributed to maintaining more ribozymes in the protocells. These circumstances are the high gene count, intermediate split size and low rate of fusion (Fig. 3A).

Bernstein et al. (1984) have raised the idea that protocellular sex may have been advantageous due to the recovery of a complete hypercycle (by re-acquiring accidentally missing genes of the set). This idea has ignored intracellular conflict and the possibility of the spread of selfish parasites, however. The latter aspect has been investigated (with some restrictions) by (Santos et al., 2003) without assuming internal hypercyclic organization, due to the finding that at high mutations rates the stochastic corrector model (group selection for balanced composition of early replicators) outperforms hypercycles (Zintzaras et al., 2002). In the present study we have investigated occasional fusion-fission episodes at different frequencies with SCM models assuming different numbers of essential genes, also including the dosage effect on cell fitness of the ribozymes. We find that the advantage of sex increases with the number of unlinked, essential replicator species, because the assortment load increases with this genome size in the first place. For the same mutation rate per gene the mutational load is also bound to increase with the number of genes: the specific beneficial effects of sex on the assortment and mutational load warrant further study. Larger protocells (with higher split size) lose the advantage of sex due the efficient horizontal spread of parasites. Sex is truly beneficial if only bad protocells (those with a markedly unbalanced gene set) can fuse and good ones enjoy celibacy with their close-to-optimal compositions being untouchable. This seems to be a robust conclusion.

In the light of the foregoing, one can justifiably ask about the mechanistic performance by protocells adhering to advantageous sex habit rule. This issue requires further investigation, but it is perhaps fair to assume that an unbalanced metabolic gene set results in an unbalanced metabolism, which in turn could upset the balanced growth of the enclosing membrane. Protocells with a stressed membrane may easily have a higher propensity to fuse—incidentally, this is a prediction for future experimental research (cf. Chernomordik and Kozlov, 2008; Malinin and Lentz, 2004; Marrink and Mark, 2003). One can envisage a process whereby metabolically (and thus, likely, osmotically) stressed protocells fuse, metabolism recovers, stress fades and fission ensues. One of the plausible causes of metabolic stress is parasites. So, if the previous reasoning is correct, parasites might have favoured their own spread indirectly through the membrane-stressing mechanism. A more direct involvement of parasites in eliciting sexual fusion cannot be ruled out, however. Hickey (1982) and Rose (1983) have raised the possibility that sex may have originated by the action of selfish elements that could induce it, even at considerable fitness costs to the cells. In our context this would be a selfish RNA (maybe even a ribozyme) that would replicate fast and would somehow elicit protocell fusion, e.g. by affecting the protocell membrane directly. It is true that specific RNA is able to bind to (Janas et al. 2006) and disrupt (Vlassov et al., 2001) membranes, although a replicating RNA “catalysing sex” is still a bit remote from these examples. Should such a possibility feasible, it would readily results in potential conflict since optimum fusion frequency could well be different for the selfish RNA and the protocell: this problem warrants further study. A further, more advanced mechanism could be fusion triggered by the regulated appearance of a membrane-anchored RNA (Stengel et al., 2007). Noteworthy in this regard is that selected RNA complexes can bind and disrupt phospholipid bilayers (Vlassov et al., 2001), that a passive membrane transporter for tryptophan built of RNA has been selected (Janas et al., 2004), and that RNAs can specifically bind to ordered phospholipid bilayers (Janas et al., 2006). Further work should determine the best lubricants for protocellular sex.

Acknowledgements

We are grateful to Harold P. de Vladar for helpful comments and ideas improving the manuscript. The authors acknowledge financial support from the Hungarian Research Fund (NKFI K119347 and K124438), the GINOP-2.3.2-15-2016-00057 and the Volkswagen Stiftung initiative “Leben? – Ein neuer Blick der Naturwissenschaften auf die grundlegenden Prinzipien des Lebens” under project “A unified model of recombination in life”.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jtbi.2018.11.020.

References

Bernstein, H., Byerly, H.C., Hopf, F.A., Michod, R.E., 1984. Origin of sex. J. Theor. Biol. 110, 323–351. doi:10.1016/S0022-5193(84)80178-2.

Boerlijst, M.C., Hogeweg, P., 1991. Spiral wave structure in pre-biotic evolution: hypercycles stable against parasites. Physica D 48, 17–28. doi:10.1016/0167-2789(91)90049-F.

Chernomordik, L.V., Kozlov, M.M., 2008. Mechanics of membrane fusion. Nat. Struct. Mol. Biol. 15, 675–683. doi:10.1038/nsmb.1455.

Czárán, T., Könnyű, B., Szathmáry, E., 2015. Metabolically coupled replicator systems: overview of an RNA-world model concept of prebiotic evolution on mineral surfaces. J. Theor. Biol. 381, 39–54. doi:10.1016/j.jtbi.2015.06.002.

Eigen, M., 1971. Selforganization of matter and the evolution of biological macromolecules. Naturwissenscaften 10, 465–523.

Grey, D., Hutson, V., Szathmáry, E., 1995. A re-examination of the stochastic corrector model. Proc. R. Soc. Lond. Ser. B 262, 29–35. doi:10.1098/rspb.1995.0172.

Hickey, D.A., 1982. Selfish DNA: a sexually-transmitted nuclear parasite. Genetics 101, 519–531.

Hubai, A.G., Kun, Á., 2016. Maximal gene number maintainable by stochastic correction – The second error threshold. J. Theor. Biol. 405, 29–35. doi:10.1016/j.jtbi.2016.02.007.

Janas, T., Janas, T., Yarus, M., 2004. A membrane transporter for tryptophan composed of RNA. RNA 10, 1541–1549. doi:10.1261/rna.7112704.

Janas, T., Janas, T., Yarus, M., 2006. Specific RNA binding to ordered phospholipid bilayers. Nucleic Acids Res. 34, 2128–2136. doi:10.1093/nar/gkl220.

Joyce, G.F., 2002. The antiquity of RNA-based evolution. Nature 418, 214–220.

Koch, A.L., 1984. Evolution vs the number of gene copies per primitive cell. J. Mol. Biol. 20, 71–76.

Könnyű, B., Czárán, T., 2013. Spatial aspects of prebiotic replicator coexistence and community stability in a surface-bound RNA world model. BMC Evol. Biol. 13, 204. doi:10.1186/1471-2148-13-204.

Kun, Á., Santos, M., Szathmáry, E., 2005. Real ribozymes suggest a relaxed error threshold. Nat. Genet. 37, 1008–1011. doi:10.1038/ng1621.

Kun, Á., Szilágyi, A., Könnyű, B., Boza, G., Zachár, I., Szathmáry, E., 2015. The dynamics of the RNA world: insights and challenges. Ann. N Y Acad. Sci. 1341, 75–95. doi:10.1111/nyas.12700.

Leu, K., Kervio, E., Obermayer, B., Turk-MacLeod, R.M., Yuan, C., Luevano, J.-M., Chen, E., Gerland, U., Richert, C., Chen, I.A., 2012. Cascade of reduced speed and accuracy after errors in enzyme-free copying of nucleic acid sequences. JACS 135, 354–366. doi:10.1021/ja3095558.

Malinin, V.S., Lentz, B.R., 2004. Energetics of vesicle fusion intermediates: comparison of calculations with observed effects of osmotic and curvature stresses. Biophys. J. 86, 2951–2964. doi:10.1016/S0006-3495(04)74346-5.

Marrink, S.J., Mark, A.E., 2003. The mechanism of vesicle fusion as revealed by molecular dynamics simulations. JACS 125, 11144–11145. doi:10.1021/ja036138+.

Matsumura, S., Kun, Á., Ryckelynck, M., Coldren, F., Szilágyi, A., Jossinet, F., Rick, C., Nghe, P., Szathmáry, E., Griffiths, A.D., 2016. Transient compartmentalization of RNA replicators prevents extinction due to parasites. Science 354, 1293–1296. doi:10.1126/science.aag1582.

Maynard Smith, J., 1978. The Evolution of Sex. Cambridge University Press.

Maynard Smith, J., Szathmáry, E., 1993. The origin of the chromosome I. Selection for linkage. J. Theor. Biol. 164, 437–446. doi:10.1006/jtbi.1993.1165.

Maynard Smith, J., Szathmáry, E., 1995. The Major Transition in Evolution. W.H. Freeman, Oxford, UK.

Michod, R., 1999. Darwinian Dynamics: Evolutionary Transition in Fitness and Individuality. Princeton University Press.

Niesert, U., Harnasch, D., Bresch, C., 1981. Origin of life between Scylla and Charybdis. J. Mol. Evol. 17, 348–353. doi:10.1007/BF01734356.

Poole, A.M., 2009. Horizontal gene transfer and the earliest stages of the evolution of life. Res. Microbiol. 160, 473–480. https://doi.org/10.1016/j.resmic.2009.07.009.

Queller, D.C., 1997. Cooperators since life began. Q. Rev. Biol. 72, 184–188.

Rajamani, S., Ichida, J.K., Antal, T., Treco, D.A., Leu, K., Nowak, M.A., Szostak, J.W., Chen, I.A., 2010. Effect of stalling after mismatches on the error catastrophe in nonenzymatic nucleic acid replication. JACS 132, 5880–5885. doi:10.1021/ja100780p.

Rasmussen, S., Bedau, M.A., Chen, L., Deamer, D., Krakauer, D.C., Packard, N.H., Stadler, P.F. (Eds.), 2008. Protocells: Bridging Nonliving and Living Matter. MIT Press, Cambridge.

Rose, M.R., 1983. The contagion mechanism for the origin of sex. J. Theor. Biol. 101, 137–146.

Santos, M., Zintzaras, E., Szathmáry, E., 2003. Origin of sex revisited. Origins Life Evol. Bioshpere 33, 405–432.

Santos, M., Zintzaras, E., Szathmáry, E., 2004. Recombination in primeval genomes: a step forward but still a long leap from maintaining a sizeable genome. J. Mol. Evol. 59, 507–519.

Scheuring, I., Czárán, T., Szabó, P., Károlyi, G., Toroczkai, Z., 2003. Spatial models of prebiotic evolution: soup before pizza? Origins Life Evol. Biosphere 33, 319–355. doi:10.1023/a:1025742505324.

Schrum, J.P., Zhu, T.F., Szostak, J.W., 2010. The origins of cellular life. Cold Spring Harb. Perspect. Biol. 2, a002212. doi:10.1101/cshperspect.a002212.

Stengel, G., Zahn, R., Höök, F., 2007. DNA-induced programmable fusion of phospholipid vesicles. JACS 129, 9584–9585. doi:10.1021/ja073200k.

Szathmáry, E., 1989. The integration of the earliest genetic information. Trends Ecol. Evol. 4, 200–204. doi:10.1016/0169-5347(89)90073-6.

Szathmáry, E., 2015. Toward major evolutionary transitions theory 2.0. PNAS 112, 10104–10111. doi:10.1073/pnas.1421398112.

Szathmáry, E., Demeter, L., 1987. Group selection of early replicators and the origin of life. J. Theor. Biol. 128, 463–486.

Szathmáry, E., Maynard Smith, J., 1995. The major evolutionary transitions. Nature 374, 227–232. doi:10.1038/374227a0.

Szathmáry, E., Santos, M., Fernando, C., 2005. Evolutionary potential and requirements for minimal protocells. Top. Curr. Chem. 259, 167–211. doi:10.1007/tcc001.

Szilágyi, A., Kun, Á., Szathmáry, E., 2014. Local neutral networks help maintain inaccurately replicating ribozymes. PLoS One 9, e109987. doi:10.1371/journal.pone.0109987.

Szilágyi, A., Zachar, I., Scheuring, I., Kun, Á., Könnyű, B., Czárán, T., 2017. Ecology and evolution in the RNA world: dynamics and stability of prebiotic replicator systems. Life 7, 48. doi:10.3390/life7040048.

Takeuchi, N., Hogeweg, P., 2012. Evolutionary dynamics of RNA-like replicator systems: a bioinformatic approach to the origin of life. Phys. Life Rev. 9, 219–263. doi:10.1016/j.plrev.2012.06.001.

Takeuchi, N., Poorthuis, P.H., Hogeweg, P., 2005. Phenotypic error threshold; additivity and epistasis in RNA evolution. BMC Evol. Biol. 5, 9. doi:10.1186/1471-2148-5-9.

Vetsigian, K., Woese, C., Goldenfeld, N., 2006. Collective evolution and the genetic code. PNAS 103, 10696–10701. doi:10.1073/pnas.0603780103.

Vlassov, A., Khvorova, A., Yarus, M., 2001. Binding and disruption of phospholipid bilayers by supramolecular RNA complexes. PNAS 98, 7706–7711. doi:10.1073/pnas.141041098.

Vogan, A.A., Higgs, P.G., 2011. The advantages and disadvantages of horizontal gene transfer and the emergence of the first species. Biol. Direct 6, 1. doi:10.1186/1745-6150-6-1.

Yarus, M., 2011. Life from an RNA World: The Ancestor Within. Harvard University Press, Harvard, USA.

Zintzaras, E., Santos, M., Szathmáry, E., 2002. “Living” under the challenge of information decay: the stochastic corrector model vs. hypercycles. J. Theor. Biol. 217, 167–181. doi:10.1006/jtbi.2002.3026.